PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04N 5/272, 5/262

A1

(11) International Publication Number: WO 97/03517

(43) International Publication Date: 30 January 1997 (30.01.97)

(21) International Application Number: PCT/GB96/01682

(22) International Filing Date: 15 July 1996 (15.07.96)

(30) Priority Data:
9514313.7 13 July 1995 (13.07.95) GB

(71)(72) Applicant and Inventor: BEATTIE, Robert, John [GB/GB]; 47 Stewartfield Drive, Eastkilbride G74 4UA (GB).

(74) Agent: MURGITROYD & COMPANY; 373 Scotland Street, Glasgow G5 8QA (GB).

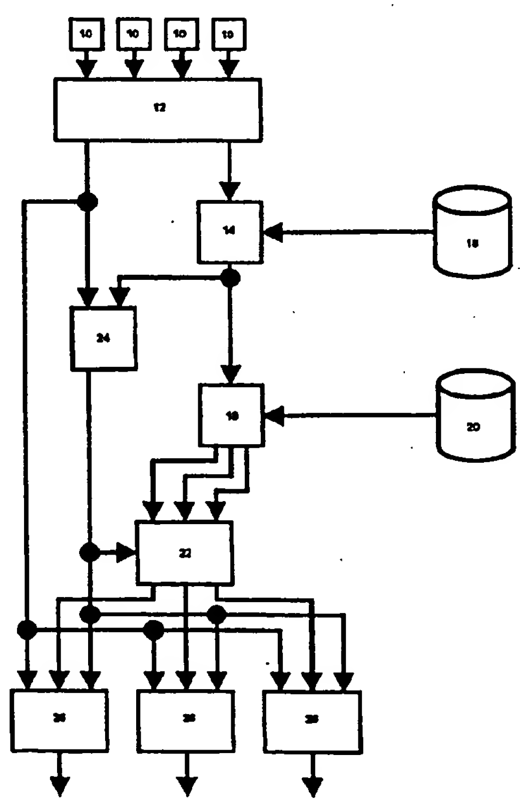
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND APPARATUS FOR PRODUCING COMPOSITE VIDEO IMAGES

(57) Abstract

A system for automatically generating and adding secondary video images (such as advertising material) to primary video images or real world scenes (such as a live sports event) in such a way that the secondary image appears to be physically present in the scene represented by the primary image when the composite image is viewed subsequently. A "live" image from one of a number of cameras (10) is selected by an editing desk (12) for transmission. Prior to transmission, a secondary image is selected from a database (20) for inclusion in the final image, such that it appears superimposed on a physical target space in the first image. The selected image is transformed in terms of size, shape, orientation and lighting effects before being combined with the primary image. The transformation is based on a computed "expected image", which is derived from a computer model (16) of the environment containing the first image (such as a sports arena) and data transmitted from the camera regarding its location, orientation, focal length, etc. The expected image is matched with the first image in a matching module (24) to refine the alignment of the computed target space with the actual target space, and to identify lighting variations and foreground objects in the first image and apply these to the second image as seen in the final composite image. Multiple composite images may be generated including different secondary images so that, for example, different advertisements can be included in different composite images for transmission to different audiences.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

"Methods and Apparatus for Producing Composite Video Images"

The present invention relates to a system for automatically generating and adding secondary images to primary images of real world scenes in such a way that the secondary image appears to be physically present in the scene represented by the primary image when the composite image is viewed subsequently.

It is particularly envisaged that the invention be applied to the presentation of advertising material (secondary images) within primary images including, but not limited to, television broadcasts, video recordings, cable television programmes and films. It is applicable to all video/TV formats, including analogue and digital video, PAL, NTSC, SECAM and HDTV. This type of advertising is particularly applicable to, but is not limited to, live broadcasts of sports events, programmes of highlights of sports events, videos of sports events, live broadcasts of important state events, television broadcasts of "pop" concerts etc.

Prior practice relating to the placement of

advertisements within scenes represented in TV/video images includes:

physical advertising hoardings which can be placed at appropriate places in a scene or venue such that they sometimes appear in the images; such hoardings can be either simple printed signs or electromechanical devices allowing the display of several fixed advertisements consecutively;

advertisements which are placed directly onto surfaces within the scene, for example, by being painted onto the outfield at a cricket match, or by being placed on players' clothes or by being painted onto racing car bodies;

small fixed advertisements, for example, company logos, which are simply superimposed on the image of the scene.

These methods have the following disadvantages:

each physical advertising hoarding can present, at most, a few static images; it cannot be substantially varied during the event, nor can its image be changed after the event other than by a painstaking manual process of editing individual images;

advertisements made, for example, on playing surfaces or on participants clothing, have to be relatively discreet otherwise they intrude too much into the event itself;

fixed advertisements, such as company logos, superimposed on the image, look artificial and intrusive since they are obviously not part of the scene being viewed.

The present invention concerns a system whereby secondary images, such as advertising material, can be combined electronically with, for example, a live action video sequence in such a manner that the

secondary image appears in the final composite image as a natural part of the original scene. For example, the secondary image may appear to be located on a hoarding, while the hoarding in the original scene contains different material or is blank. This allows, for example, different advertising material to be incorporated into the scene to suit different broadcast audiences.

Numerous systems exist for combining video images for various purposes. The prior art in this field includes the use of "colour keying" (also known an "chroma keying") in which a foreground object, such as a weather forecaster, is in front of a uniform background of a single "key" colour. A second video source provides another signal, such as a weather map. The two video signals are mixed together so that the second video signal replaces all parts of the first video signal which have the key colour. A similar approach is employed in "pattern-keying". Alternatively, of course, individual frames of the primary image could be edited manually to include the secondary image.

It has previously been proposed to use video systems of this general type to insert advertising material into video images, one example being disclosed in WO93/02524. WO93/06691 discloses a system having similar capabilities.

Colour keying works well in very restricted circumstances where the constituent images can be closely controlled, such as in weather forecasting or pre-recorded studio productions. However, it does not work in the general case where it is desired to mix unrestricted background images in parts of unrestricted primary images. The same applies generally to pattern-

keying systems. Replacing physical advertising signs by manually editing series of images is not feasible for live broadcasts and is extremely costly even for use with recorded programmes.

Existing systems such as these are not well suited for the purposes of the present invention. Even where prior proposals relate specifically to the insertion of advertising material in video images, such proposals have not addressed one or more issues such as coping with foreground objects or with lighting effects or with multiple cameras.

In accordance with a first aspect of the present invention there is provided a method of modifying a first video image of a real world scene to include a second video image, such that said second image appears to be superimposed on the surface of an object appearing within said first image, wherein said second image is derived by transforming a preliminary second image to match the size, shape and orientation of said surface as seen in said first image and said second image is combined with said first image to produce a composite final image;

said method including:

a preliminary step of constructing a three-dimensional computer model of the environment containing the real world scene, said model including at least one target space within said environment upon which said second image is to be superimposed;

generating camera data defining at least the location, orientation and focal length of a camera generating said first image; and

transforming the preliminary second image on the basis of said model and said camera data so as to match said target space as seen in the first image, prior to

combining said first image and said second image.

In accordance with a second aspect of the invention there is provided apparatus for generating a composite video image comprising a combination of a first video image of a real world scene and a second video image, such that said second image appears to be superimposed on the surface of an object appearing within said first image, including:

at least one camera for generating said first image;

means for generating said second image by transforming a preliminary second image to match the size, shape and orientation of said surface as seen in said first image; and

means for combining said second image with said first image to produce a composite final image;

said apparatus including:

means for storing a three-dimensional computer model of the environment containing the real world scene, said model including at least one target space within said environment upon which said second image is to be superimposed;

means for generating camera data defining at least the location, orientation and focal length of a camera generating said first image; and

means for transforming the preliminary second image on the basis of said model and said camera data so as to match said target space as seen in the first image, prior to combining said first image and said second image.

Further aspects and preferred features of the invention are defined in the Claims appended hereto.

Embodiments of the invention will now be described, by

way of example only, with reference to the accompanying drawing, which is a schematic block diagram of a system embodying the invention.

The overall scheme of the invention is illustrated in the drawing. One or more cameras 10 are deployed to provide video coverage of an event in a venue, such as a sporting arena (not shown). The following discussion relates particularly to "live" coverage, but it will be understood that the invention is equally applicable to processing pre-recorded video images and associated data.

Each of the cameras 10 is augmented by the addition of a hardware module (not shown) adapted to generate signals containing additional data about the camera, including position and viewing direction in three dimensions, and lens focal length. A wide variety of known devices may be used for providing data about the orientation of a camera (e.g. inclinometers, accelerometers, rotary encoders etc.), as will be readily apparent to those of ordinary skill in the art.

The video signal from each camera 10 in operation at a particular event is passed to an editing desk 12 as normal, where the signal to be transmitted is selected from among the signals from the various cameras.

The additional camera data is passed to a modelling module (computer) 14 which has access to a predefined, digital 3-d model of the venue 16. The venue model 16 contains representations of all aspects of the venue which are significant for operation of the system, typically including the camera positions and the locations, shapes and sizes of prominent venue features and all "target spaces" onto which secondary images are

to be superimposed by the system, such as physical advertising hoardings.

The modelling module 14 uses the camera location, orientation and focal length data to compute an approximation of the image expected from the camera 10 based on transformed versions of items forming part of the model 16 which are visible in the camera's current view.

The modelling module 14 also calculates a pose vector relative to the camera view vector for each of the target spaces visible in the image. Target spaces into which the system is required to insert secondary images are referred to herein as "designated targets".

The additional camera data is also passed to the secondary image generation module 18 which generates a preliminary secondary image for each designated target in the primary image. A library of secondary images is suitably stored in a secondary image database 20, accessible by the secondary image generation module 18.

The pose of each of the designated targets, derived from the "expected view" calculated by the modelling module 14, is fed into a transformation module 22 together with the preliminary secondary images. The preliminary secondary images are transformed by the transformation module 22 so that they have the correct perspective appearance (size, shape and orientation) to match the corresponding target space as viewed by the camera 10.

The original video image and the expected image calculated from the 3-d model 16 are both also passed to a matching module 24. The matching module 24

effectively superimposes the calculated expected image over the actual image as a basis for matching the two. It identifies as many as possible of the corners and edges of the target spaces corresponding to the designated targets and any other items of the venue model 16 present in the expected image. It uses these matches to refine the transformational match of the expected image to the actual image. Finally, the matcher extracts any foreground objects and lighting effects from the image areas of the designated targets.

The original primary image from the editing desk 12, the transformed secondary image and the output data from the matching module 24 are passed to one or more output modules 26 where they are combined to produce a final composite video output, in which the primary and secondary images are combined. There may be multiple output modules 26, each inserting different secondary images into the same primary images.

Obviously, for live transmission, this whole procedure has to happen in real time. Fortunately, the state of modern computing and image processing technology is such that the necessary hardware is not particularly expensive.

Each of the modules mentioned above is described in more detail below.

Camera Augmentation

Each camera is equipped with a device which continuously transmits additional camera data to the central station. This camera data could either be transmitted via a separate means such as additional cables or radio links, or could be incorporated into

the hidden parts of the video signal in the same way as teletext information. Methods and means for transmitting such data are well known.

This camera data typically includes some or all of:

a camera identifier;
the camera position;
the camera orientation;
the lens focal length;
the lens focusing distance;
the camera aperture.

The camera identifier is a string of characters which uniquely identifies each camera in use. The camera position is a set of three coordinate values giving the position of the camera in the coordinate system in use in the 3-d venue model. The camera orientation is another set of three values, defining the direction in which the camera is pointing. For example, this could be made up of three angles defining the camera viewing direction in the coordinate system used to define the camera position. The coordinate system used is not critical as long as all the cameras in use at a particular event supply the camera data in a way which is understood by the modelling and transformation modules.

Since most cameras are fitted with zoom lenses, the lens focal length is required to define the scene for the purposes of secondary image transformation. The lens focusing distance and camera aperture are also required to define the scene for the purposes of transforming the secondary image in terms of which parts of the scene are in focus.

The additional devices with which each camera is

equipped may depend on the role of the camera. For example, a particular camera may be fixed in position but adjustable in orientation. In this case, a calibration procedure may be used which results in an operator entering the camera's position into the device before the event starts. The orientation would be determined continuously by the device as would the focal length, focusing distance and aperture.

The Venue Model

Key elements at the venue are represented within the general 3-d venue model 16.

The model may be based on a normal orthogonal 3-d coordinate system. The coordinate system origin used at a particular venue may be global or local in nature. For example, if the venue is a soccer stadium, it may be convenient to take the centre spot as the origin and to take the half-way line to define one axis direction, with an imaginary line running down the centre of the pitch defining a second axis direction. The third axis would then be a vertical line through the centre spot.

Each relevant permanent item of the venue is represented within the model in a way which encapsulates the item's important features for the purposes of the present system. Again, in the example of the soccer stadium, this could include:

the playing surface, represented as a planar surface with particular surface markings and a particular texture;

goalposts, represented as a solid object, for example, as the intersection of several cylindrical objects, having specific surface properties, e.g. white colour;

goal nets, which may be represented as an intersection of curvilinear objects with specific surface properties and having the property of flexibility;

advertising hoardings, which, in the simplest case, are represented as planar surfaces with complex surface properties, i.e. the physical advertisement (it is preferable that the surface properties are stored using a scale-invariant representation in order to simplify the matching process);

prominent permanent venue features: it is useful to the matching process if prominent features are included in the venue model; these may be stored as solid objects with surface properties (for example, if a grandstand contains a series of vertical pillars, then these could be used in the matching process to improve the accuracy of the process).

The methods and means for generating and using 3-d models, such as the venue model described above, and for determining the positions of objects within such models are all well known from other applications such as virtual reality modelling.

Overall Signal Processing

The object of the signal processing performed by the system is to identify the position of the designated targets in the current image, to extract any foreground objects and lighting effects relevant to the designated targets, then to generate secondary images and insert them into the current primary image in place of the designated targets such that they look completely natural. The signal processing takes place in the following stages.

1. Use the camera data in conjunction with the venue model to generate an expected image incorporating all the objects in the venue model which are expected to be seen in the actual image and to calculate the pose of each of the visible designated targets relative to the camera (modelling module 14).

2. Identify as many as possible of the expected objects in the actual image (matching module 24).

3. Use the individual item matches to refine the view details of the expected image (matching module).

4. Project the borders of the designated targets onto the real image and refine the border positions, where appropriate with reference to edges and corners in the actual image (matching module 24).

5. Match the expected designated target image to the corresponding region in the actual image, the match to be performed separately in colour space and intensity space. Any missing regions in the colour space match are assumed to be foreground objects. The bounding subregion of the target region is extracted and stored. The stored region includes colour and intensity information. Any mismatch regions occurring in intensity space only, e.g. shadows, which are not part of foreground objects are extracted and stored as intensity variations (matching module 24).

6. Store the outcome of the matching process for use in matching the next frame.

7. Transform the scale-invariant designated target model to fit the best estimate bounding region (transform module 22).

8. Reassemble as many outgoing video signals as required by inserting the transformed secondary images into the original primary image and then reinserting foreground objects and lighting effects (output module).

Matching Module

The matching module 24 has several related functions.

The matcher first compares the expected view with the actual image to match corners and edges of items in the expected view with corresponding corners and edges in the actual image. This is greatly simplified by the fact that the expected ~~image should be~~ very close to the same view of the scene as the actual image. The object of this phase of matching is to correlate regions of the actual image with designated targets in the expected image. Corners are particularly beneficial in this part of the process since a corner match provides two constraints on the overall transformation whilst an edge match provides only one. Since the colour of the objects in the expected image is known from their representation in the venue model, this provides a further important clue in the matching process. When as many as possible of the corners and edges of the objects in the expected image have been matched to corners and edges in the actual image, a consistency check is carried out and any individual matches which are inconsistent with the overall transformation are rejected. Matching corners and edges in this way is a method well established in machine vision applications.

The outcome of the first phase of matching is a detailed mapping of the expected image onto the actual image. The second stage of matching is to deal with each designated target in turn to identify its exact boundary in the image and any foreground objects or lighting effects affecting the appearance of the corresponding physical object or area in the original image. This is done by using the corner and edge

matches and interpolating any missing sections of the boundary of the original object/area using the projected boundary of the designated target. For example, if the designated target is a rectangular advertising hoarding, then as long as sufficient segments of the boundary of the hoarding are identified, the position of the remaining segments can be calculated using the known segments and the known shape and size of the hoarding together with the known transformation into the image.

The final stage of the matching process involves identifying foreground objects and lighting effects within the region of each designated target. This is based on transforming the scale invariant representation of the designated target in the venue model such that it fits exactly the bounding region of the corresponding ad in the original image. A match in colour space is then carried out within the bounding region to identify sections of the image which do not match the corresponding sections of the transformed model. These non-matching sections are taken to be foreground objects and these parts of the image are extracted and stored to be superimposed on top of the transformed secondary image in the final composite image. A match in intensity space is also carried out to identify intensity variations which are not part of the original object/area. These are considered to be lighting effects and an intensity transformation is used to extract these and keep them for later use in transforming the secondary image.

Hence, the output from the matching process includes:

the exact image boundary of all the designated targets;

foreground objects in any of these regions; and

lighting effects in any of these regions.

Secondary Image Generation Module

One of the major advantages of using electronically generated secondary images rather than physical signs is in the extra scope for controlling the choice, positioning and content of the secondary image, e.g. an advertising message.

Generation of the secondary images uses a database 20 of secondary image material. In addition to the actual secondary images, stored as scale-invariant representations, this database may include information such as:

the percentage of the available advertising space-time has been booked by each advertiser;

any preferences on which part of the event's duration and which part of the venue are to be used for each advertiser;

associations of particular secondary images with potential occurrences in the event being covered.

Another strength of the use of electronically integrated secondary images is the ability to generate different video outputs for different customers. Hence, in an international event, different advertising material could be inserted into the video signal going to different countries. For example, say the USA is playing China at basketball. Most Americans don't read Chinese and most Chinese don't read English. So the transmission to China would include only advertisements in Chinese, while the broadcast in the USA would include only english language advertisements.

Generating a particular advertisement for display in

the present system may take place in the following stages:

choose the company whose advertisement will be displayed;

choose which of the selected company's advertisements is appropriate for the current context;

transform the stored representation of the selected advertisement to match the available region of the image.

For the first stage of this process, the selection of the advertiser, the destination of the video signal concerned is first determined. This indexes the advertisers for the output module 26 corresponding to that destination. Next, a check is made to see how much advertising time each advertiser has had during the event so far relative to how much they have booked. The advertiser is selected on this basis, taking account of advertiser preferences such as location and timing.

The next stage, the selection of one advertisement from a set supplied by the advertiser to replace a designated target in the original image, is based on factors including:

the size of the space available;

the location of the designated target;

the phase of the event;

any notable occurrences during the event.

For example, an advertiser may choose to supply some advertisements containing a lot of detail and some which are very simple. If the space available is large, perhaps because the camera concerned is showing a close up of a soccer player about to take a corner and the advertising space available fills a large part

of the image, then it may be appropriate to fit a more detailed advertisement where the details will be visible. At the other extreme, if a particular camera is showing a long view, then it may be better to select a very simple advertisement with strong graphics so that the advertisement is legible on the screen.

Note also that the selection of advertisements can be influenced by what has happened in the event. For example, say a particular player, X, has just scored a goal. Then an advertiser who manufactures drink, Y, may want to display something to the effect that "X drinks Y". To meet this need the system has the capability to store advertisements which are only active (i.e. available for selection) when a particular event has taken place. Additionally, these advertisements can have place holders where the name of a participant or some other details can be entered when the ad is made active. This could be useful if drinks advertiser Y has a contract with a whole team. Then when any team member does something exceptional, that team member's name, or other designation, could be inserted into the advertisement.

Note also that there is no restriction on advertisements being static. As long as the advertisement still looked as though it was part of the event, it could be completely dynamic. For example, an advertising video could be inserted into a suitable designated target. One particular case might be where the venue concerned has a large playback screen, such as at many cricket and athletics events. The screen would be used to show replays of the event to the spectators present, but it could also be a designated target for the present system. Such a screen would then be a good candidate for showing video advertising

material.

A further aspect of the process of secondary image generation relates to how to change images. Clearly, if a camera is panning, then different secondary images can be included as different parts of the venue come into the image. Note that it is important to record which secondary image is being displayed on which designated target, since a cut from one camera to another should not cause the secondary image to change if the two cameras are capturing the same designated target. It can also occur that one camera will be used for a particularly long time and it and it may be desirable to change the secondary images in the composite image part way through the shot. This is accomplished by simulating the change of a physical ad. For example, there are physical advertising hoardings available which are able to show more than one ad, either by rotating a strip containing the ads or by rotating some triangular segments, each of whose faces contains portions of different ads. To change a secondary image while it is in shot, the secondary image generation process may simulate the operation of a physical hoarding, for example, by appearing to rotate segments of a hoarding to switch from one ad to the next.

<u>Transform Module</u>

The pose of the physical advertising space relative to the camera concerned is known from the additional camera data and the 3-d venue model 16. Hence, transforming the scale-invariant representation of the chosen secondary image into a 2-d image region with the correct perspective appearance is a straightforward task. In addition to the pose being correct, the

secondary image has to fit the target space exactly. The region bounding the space is supplied by the matching process. Hence, transforming the ad involves:

using the additional camera data and 3-d venue model 16 to calculate the perspective appearance of the secondary image (this is done in the modelling module 14);

using the matching information to scale the secondary image to fit the space available.

The secondary image is now ready to be dropped into the original video image.

Output Module

One output module 26 is required for each outgoing video signal. Hence, if the final of the World Cup is being transmitted to 100 countries which have been split into 10 areas for advertising, then ten output modules would be required.

The output module 26 takes one set of secondary images and inserts them into the original primary image. It then takes the foreground object and lighting effects generated by the matching process and reintegrates them. In the case of the foreground objects, this requires parts of the inserted secondary images to be overwritten with the foreground objects. In the case of lighting effects, such as shadows, the image segments containing the secondary image must be modified such that the secondary image looks as if it is subjected to the same lighting effects as the corresponding part of the original scene. This is done by separating out the colour and intensity information and modifying them appropriately. Methods for doing this are well known in the field of computer graphics.

Use of the present invention has many benefits for advertisers, particularly at large international events. Some of these benefits are as follows:

different advertisements can be shown in different countries or regions thereby improving targeting and making sure that the advertising regulations of individual countries, e.g. with respect to alcohol and tobacco, are not violated;

- each advertiser can be guaranteed a percentage of the total exposure;

the detail of the advertisements can be adjusted automatically based on their size in the TV image to improve their legibility and impact;

there may be much greater creative scope in the design of the advertisements;

by recording some extra information with the individual camera video signals, different advertisements can be used in subsequent use of the original footage: for example, different advertisements could be used in programmes of highlights than in live broadcasts, and different advertisements again could be used in subsequent video products.

Systems for replacing parts of video images with parts of other images such that the replacement parts appear to be a natural part of the original image are known in the prior art. However, the systems described in the prior art have serious limitations which are overcome by the present invention.

One area of the prior art is based on colour or chroma keying. This depends on being able to control the colour of everything in the image and is not practical as a general purpose system.

Another area of prior art involves a human operator

manually selecting the areas to be replaced and performing various functions to deal with foreground objects and lighting effects. This method is very time consuming and expensive and obviously not applicable to live broadcasts.

Another area of prior art specifies automatic replacement of an advertising logo using the pose of the identified logo to transform the virtual ad (WO93/06691). However, this method does not describe any way of dealing with foreground objects or lighting effects.

The main advantages of the present invention over the prior art are considered to be:

augmentation of cameras and the use of a full 3-d venue model to enable generation of an expected image and reliable and fast matching of the expected image to an actual image without relying on colour keying or extensive searching or analysis of the actual image;

use of the full 3-d venue model together with the additional camera data to eliminate the need to estimate the pose of physical ads from the image data;

separation of the video signal into colour and intensity images for separate treatment of foreground objects and lighting effects;

use of corner and edge detection and matching as the basis for superimposing expected image segments over actual image segments;

use of stored scale-invariant representations of the physical designated targets to greatly simplify identification of foreground objects and lighting effects.

As a result of these improvements, the present invention is much more generally applicable than those

based on the prior art.

Improvements and modifications may be incorporated without departing from the scope of the invention.

Claims

1. A method of modifying a first video image of a real world scene to include a second video image, such that said second image appears to be superimposed on the surface of an object appearing within said first image, wherein said second image is derived by transforming a preliminary second image to match the size, shape and orientation ~~of said surface.~~ as seen in said first image and said second image is combined with said first image to produce a composite final image;

said method including:

a preliminary step of constructing a three-dimensional computer model of the environment containing the real world scene, said model including at least one target space within said environment upon which said second image is to be superimposed;

generating camera data defining at least the location, orientation and focal length of a camera generating said first image; and

transforming the preliminary second image on the basis of said model and said camera data so as to match said target space as seen in the first image, prior to combining said first image and said second image.

2. A method as claimed in Claim 1, wherein the transformation of the preliminary second image includes manipulation thereof to take account of lighting conditions in the image of the real world scene.

3. A method as claimed in Claim 2, wherein objects included in said model are matched with corresponding regions of said first image, intensity information relating to matched objects is compared with intensity information relating to said corresponding image region, regions of intensity mismatch within said

corresponding regions are identified as lighting variations and, when said second image is transformed, the intensity of portions thereof is varied on the basis of said regions of intensity mismatch so as to simulate lighting variations within the first image.

4. A method as claimed in any preceding Claim, wherein the combination of the first and second images includes manipulation thereof to take account of foreground objects in the image of the real world scene.

5. A method as claimed in Claim 4, wherein objects included in said model are matched with corresponding regions of said first image, colour information relating to matched objects is compared with colour information relating to said corresponding image region, regions of colour mismatch within said corresponding regions are identified as foreground objects and, when said first and second images are combined, said first image is retained in preference to said second image within said colour mismatch regions.

6. A method as claimed in any preceding Claim, wherein said camera data and said computer model are combined to compute a representation of the image expected from the camera.

7. A method as claimed in Claim 6, wherein features of said expected image are matched with features of said first image.

8. A method as claimed in Claim 7, wherein said matching of the expected image and the first image is used to refine the boundary of the target space within the expected image.

9. A method as claimed in Claim 8, wherein the transformation of the shape, size and orientation of the preliminary second image is based on said refined target boundary.

10. A method as claimed in Claim 7, Claim 8 or Claim 9, wherein said matching of the expected image and the first image includes comparison of colour and intensity information for the purpose of identifying foreground objects and lighting variations in said first image.

11. A method as claimed in any one of Claims 7 to 10, wherein said first image and said second image are combined on the basis of said matching of features between the expected image and the first image.

12. A method as claimed in any one of Claims 7 to 11, wherein said computer model includes scale-invariant colour representations of surface properties of said target spaces and said expected image incorporates said colour representations of said target spaces.

13. A method as claimed in any preceding Claim, wherein said first video image is a live action video image and said composite image is generated in real time.

14. A method as claimed in any preceding Claim wherein multiple second images are superimposed upon multiple target spaces.

15. A method as claimed in any preceding Claim, wherein multiple composite images are generated, each comprising the same first image combined with differing second images.

16. A method as claimed in any preceding Claim, wherein said second image is selected automatically from a plurality of images, in accordance with predetermined selection criteria.

17. A method as claimed in any preceding Claim, wherein said first image is selected from a plurality of video images generated by a plurality of cameras.

18. Apparatus for generating a composite video image comprising a combination of a first video image of a real world scene and a second video image, such that said second image appears to be superimposed on the surface of an object appearing within said first image, including:

at least one camera for generating said first image;

means for generating said second image by transforming a preliminary second image to match the size, shape and orientation of said surface as seen in said first image; and

means for combining said second image with said first image to produce a composite final image;

said apparatus including:

means for storing a three-dimensional computer model of the environment containing the real world scene, said model including at least one target space within said environment upon which said second image is to be superimposed;

means for generating camera data defining at least the location, orientation and focal length of a camera generating said first image; and

means for transforming the preliminary second image on the basis of said model and said camera data so as to match said target space as seen in the first image, prior to combining said first image and said

second image.

19. Apparatus as claimed in Claim 18, wherein the means for transforming the preliminary second image includes means for manipulating said second image to take account of lighting conditions in the first image of the real world scene.

20. Apparatus as claimed in Claim 19, means for matching objects included in said model with corresponding regions of said first image, said matching means including means for comparing intensity information relating to matched objects with intensity information relating to said corresponding image region, and means for identifying regions of intensity mismatch within said corresponding regions, and wherein said image transforming means includes means for varying the intensity of portions of said second image on the basis of said regions of intensity mismatch so as to simulate lighting variations within the first image.

21. Apparatus as claimed in any one of Claims 18 to 20, wherein the means for combining the first and second images includes means for manipulating said second image to take account of foreground objects in the image of the real world scene.

22. Apparatus as claimed in Claim 21, including means for matching objects included in said model with corresponding regions of said first image, said matching means including means for comparing colour information relating to matched objects with colour information relating to said corresponding image region, and means for identifying regions of colour mismatch within said corresponding regions, and wherein

said image combining means includes means for manipulating said second image such that, when said first and second images are combined, said first image is retained in preference to said second image within said colour mismatch regions.

23. Apparatus as claimed in any one of Claims 18 to 22, including computer modelling means adapted to compute a representation of the image expected from the camera on the basis of said camera data and said computer model.

24. Apparatus as claimed in Claim 23, including means for matching features of said expected image with features of said first image.

25. Apparatus as claimed in Claim 24, wherein said means for matching the expected image and the first image is further adapted to refine the boundary of the target space within the expected image.

26. Apparatus as claimed in Claim 25, wherein the image transformation means is adapted to effect transformation of the shape, size and orientation of the preliminary second image based on said refined target boundary.

27. Apparatus as claimed in Claim 24, Claim 25 or Claim 26, wherein said means for matching the expected image and the first image includes means for comparing colour and intensity information for the purpose of identifying foreground objects and lighting variations in said first image.

28. Apparatus as claimed in any one of Claims 24 to 27, wherein said means for combining said first image

and said second image are adapted to effect said combination on the basis of said matching of features between the expected image and the first image.

29. Apparatus as claimed in any one of Claims 24 to 28, wherein said computer model includes scale-invariant colour representations of surface properties of said target spaces and said modelling means is adapted to generate expected images incorporating said colour representations of said target spaces.

30. Apparatus as claimed in any one of Claims 18 to 29, wherein said first video image is a live action video image and the apparatus is adapted to generate said composite image in real time.

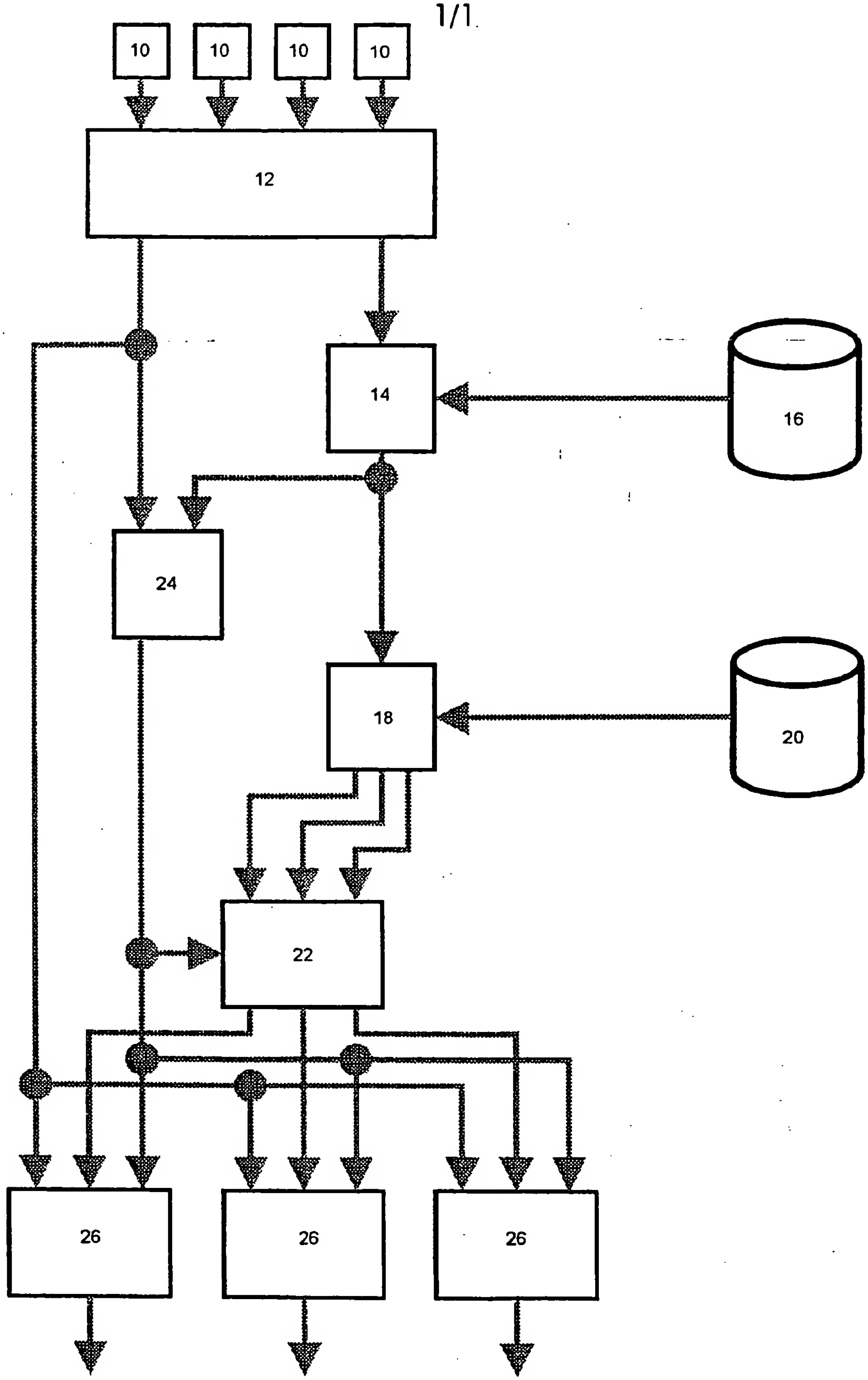
31. Apparatus as claimed in any one of Claims 18 to 30, wherein the apparatus is adapted to superimpose multiple second images upon multiple target spaces.

32. Apparatus as claimed in any one of Claims 18 to 31, including multiple output means, each of said output means being adapted to generate different composite images, each of said different composite images comprising the same first image combined with differing second images.

33. Apparatus as claimed in any one of Claims 18 to 32, including means for storing a plurality of images and means for automatically selecting said second image from said plurality of images, in accordance with predetermined selection criteria.

34. Apparatus as claimed in any one of Claims 18 to 33, wherein a plurality of cameras are connected to video editing means and said first image is selected

from a plurality of video images generated by said plurality of cameras.


10
10
10
10
12
14
16
24
18
20
22
26
26
26

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 96/01682

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04N5/272 H04N5/262

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO,A,95 10919 (ORAD INC ;SHARIR AVI (IL); TAMIR MICHAEL (IL)) 20 April 1995 see page 3, line 4 - page 11, line 16 see page 15, line 4 - page 36, line 30; figures 1-18	1-34
X	WO,A,93 06691 (SARNOFF DAVID RES CENTER) 1 April 1993 cited in the application	1,2,4,6, 7,10,11, 13,14, 17-19, 21,23, 24,27, 28,30, 31,34
A		3,5,9, 20,22,26
	see page 7, line 34 - page 21, line 2; figures 4-7	

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
7 November 1996	- 4. 12. 96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	De Paepe, W

Form PCT/ISA/210 (second sheet) (July 1992)

International Application No
PCT/GB 96/01682

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	US,A,5 491 517 (KREITMAN HAIM ET AL) 13 February 1996	1,2,4,6, 7,10,11, 13, 17-19, 21,23, 24,27, 28,30,34
P,A		3,5,9, 14,20, 22,26, 31-33
	see column 1, line 59 - column 3, line 44 see column 5, line 51 - column 6, line 2; figure 2 see column 6, line 50 - column 6, line 64; figures 1,2 see column 7, line 42 - column 8, line 54; figures 5-9 see column 12, line 59 - column 14, line 26; figures 18-21	

2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 96/01682

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9510919	20-04-95	AU-A-	6298794	04-05-95
		BR-A-	9406756	02-04-96
		CN-A-	1119481	27-03-96
		EP-A-	0683961	29-11-95
WO-A-9306691	01-04-93	AU-B-	663731	19-10-95
		AU-A-	2672092	27-04-93
		CA-A-	2119272	01-04-93
		EP-A-	0604572	06-07-94
		JP-T-	6510893	01-12-94
		US-A-	5566251	15-10-96
US-A-5491517	13-02-96	AU-A-	1933495	03-10-95
		CA-A-	2179031	21-09-95
		WO-A-	9525399	21-09-95

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.